



Concurrent Sessions



Outcome Evaluation/Outcome Monitoring

Facilitators:	David Cotton, Jeanine Ambrosio
CDC Representatives:	Charles Collins, Gary Uhl
Health Department Peers:	Roger Myrick, CA and Marcia Sass, NJ
CBO Peer:	Maija Neville

The focus of these sessions was outcome evaluation and monitoring. The group will discuss what a jurisdiction can do if they do not have adequate baseline data for doing outcome evaluation, and if it is possible to use a time series design instead of using a control/comparison group. The group also discussed TA needs, including designing outcome evaluations, identifying evaluable interventions, creating appropriate outcome measures, selecting control or comparison groups, monitoring data quality, conducting interim and final data analysis, interpreting results, and understanding the distinction between OE and OM.

This session was convened twice. For the most part, the presentations and exercises were the same, so they have not been captured twice. The discussion/input summaries from the participants, however, are documented separately for each day in order to reflect the similarities and differences in each group.

Charles Collins
CDC Representative
Opening Remarks

Charles Collins called the session to order indicating that there would be a change given that

about 48 hours earlier they'd been involved in IRB issues surrounding monitoring and evaluation. They are being told that quasi-experimental designs and experimental designs may be deemed inappropriate for use of cooperative agreement funds because these designs fall into the area of research rather than evaluation. Therefore, a decision was made to hold back on the originally scheduled agenda and have Marlene Glassman field a discussion session. While they don't have a lot of answers, CDC thought that by knowing the questions, they could work on formulating those answers so they can give participants guidance on these issues.

Marlene Glassman
CDC Representative

Marlene Glassman referred the participants to a draft letter in their packets which they included to lay out the issues. She stressed that the letters were drafts pending the approval of the CDC Procurement and Grants Office (PGO) that has jurisdiction over the use of funding. She reviewed the information in the drafts, and then fielded the discussion.

Discussion Summary:

Morning Session

- ☐ An inquiry was posed as to how CDC planned to deal with recommendations. Marlene Glassman indicated that they would review recommendations/questions posed during the meeting once they returned to their offices. In addition, they will probably create a written instrument to find out what everyone is doing. She pointed out that most of them to whom she had spoken were really engaged in outcome monitoring, which is fine. She stressed that they would have the opportunity to talk about this on a case-by-case basis.
- ☐ An inquiry was posed about when the Chapter 7 revisions would be out, and what the IRB issues meant in terms of the upcoming applications with regard to submitting their outcome evaluations. Marlene Glassman responded that Chapter 7 would be revised as soon as this issue was resolved, hopefully within a month. If it turns out that the outcome evaluation is out, then they will have a "meeting of the minds" about outcome monitoring. Then they can write about their plans for that.
- ☐ An inquiry was posed as to what the key issue was. Marlene Glassman responded that it was research design.
- ☐ A participant noted that some of them had gotten IRB approval for what they thought they were doing in the way of outcome monitoring because they're collecting confidential data on a pre-post basis from clients. Marlene Glassman clarified that it was not clear when local IRBs would come into play, but she encouraged participants to go

through their local IRBs to ensure that they are following appropriate procedures.

- ❑ An inquiry was posed as to whether there was any contemplation of changing the timeline (e.g., the evaluation is due at the end of the Cooperative Agreement in 2003) in light of this development. Marlene Glassman responded that they could not shift it under the current cooperative agreement because that is getting toward the end. If there is a delay, and it turns out to be just a delay of 4 weeks or so, they probably would not consider extension. However, if there is a delay longer than that, then they will reconsider.
- ❑ A participant suggested that CDC check with each grantee regarding whether they have multiple project assurances or federal wide assurance, because issues with multiple project assurances can cause extreme delays.

Afternoon Session

- ❑ One participant wondered how long it was going to take to find out the outcome of this situation. If it's not going to take a long time, he would be inclined not to go back and even tell anybody that things have stopped because it takes so long to get things through his own process. Marlene Glassman responded that she didn't think it would be more than 2 to 4 weeks to reach resolution. She noted that they had one determination already, and she explained the position of the PGO office to this group as well.
- ❑ An inquiry was posed as to what would happen if a group was in the middle of an outcome, quasi-experimental evaluation. Marlene Glassman said that if they had local IRB which made a determination between whether it was formative evaluation or research, this could help answer the question. She referred to the draft which instructs those with programs in progress not to continue to enroll clients until a decision is made. She said that they all shared the same concerns – that ultimately someone might not get needed services due to lack of ability to carry out interventions. So, they're working as diligently as possible to resolve the situation. She stressed that groups which were not already in the middle of an evaluation should hold off on starting one until a decision was reached. She suggested that anyone doing experimental or quasi-experimental designs should stop enrolling clients. If it's going to be an enormous problem, she suggested that they call their Project Officer for one-on-one counseling about what to do.
- ❑ An inquiry was posed as to whether they could just stop the comparison group but continue with the work for the treatment group, because that would be the definition of outcome monitoring. Some expressed concern that if they stopped the comparison group, they'd then be told in a month or two that it would be okay to go ahead. Then they will have cut off many people who could have potentially provided valuable

information to the outcome of the intervention. Charles Collins responded that the Grants Office has the ability to interpret the Cooperative Agreements. At the same time, the Human Research Office is telling them the IRB part, saying they have to shut down any kind of experimental work done by the states with these moneys. Simultaneously, the Administration is saying that all of this went through CDC clearance already. Marlene Glassman stressed that for those who were underway and who had contracts, they will *try* to help seek other types of funding streams so that the work won't be interrupted.

David Cotton, Facilitator

Overview of the Evaluation Pyramid

David Cotton gave an overview of some of the underlying principles around the relationships between evaluation activities. Referring to the diagram of the pyramid, he explained that it reflected the relationship between the different activities that go on between Community Planning, Funding, Services/Interventions, Implementing Programs, and the expected relationship with changes in risk determinants and changes in HIV transmission.

He explained the basic logic, stressing that it was not quite as linear as the diagram made it out to be, but in terms of CDC funding, the underlying logic suggests that there is a planning process in which priorities are determined – both for priority populations and for intervention strategies for most effective services to help prevent HIV. Based on that comprehensive plan of priorities, that should lead to an application to CDC for funding which corresponds to those priorities. Hopefully, there will be interventions designed that address what's asked for in the application, that funds are allocated, and those things are implemented (hopefully well and with integrity to the original design). If so, it should lead to changes in risk determinants and ultimately in HIV transmission – or at least a cumulative effect of all of those things in a particular jurisdiction.

The Guidance was designed around this logic and has evaluation components that correspond to each of these activities that are parts of the planning, implementation, and results cycle.

One of the ways to think about this is that these different activities also create a foundation on which to build evidence to support programs. The bottom of the foundation is really around the prevention priorities in the case of HIV in that if the priority setting process goes well and is agreed upon in the community, there should be priorities which serve as a foundation for a combination of both science and community input for both the population to be served and the priorities. Intervention plans would then build on those priorities. If a jurisdiction does not have good interventions which match the priorities, they've essentially lost a layer in their foundation. Similarly, if they have interventions that are well designed but they're not implemented as intended, then there is another place where a break in the chain might occur.

This is another reason that process monitoring is strongly emphasized throughout this process, because there is a critical assumption being made in any kind of intervention work that they're *actually* implementing the thing that they *think* they are. There may be many good designs and a million reasons it's not actually put into the field the same way that the designers believe that it should be working. An intervention that is not mature in that respect, also may not have the expected results. This brings up the issue of outcomes. The underlying assumption in looking at outcomes is that all of these pieces are in place in a relatively complete way.

Outcome monitoring does not answer the question of attribution. Outcome monitoring, as it's used in the Guidance, is really only looking at simple pre/post measures of certain outcomes that are the objectives of the intervention. That kind of measurement will not tell them that any changes that they do see can be attributed to that intervention. It only says that for some reason, things are moving in the right direction. What outcome monitoring does provide is a warning flag if the expected changes are not being seen. It's really an early warning system that design and plans need to be revisited.

Outcome evaluation, on the other hand, puts into play design characteristics that allow a program to rule out other sources of possible influence on the relationship between the intervention and the outcomes being seen. Outcome monitoring is very important in that it provides an early warning system, and David Cotton said he personally believed that all programs should have a provision for outcome monitoring in place because they ultimately wanted to have some initial indication about whether the hypothesized outcomes are being achieved. Outcome evaluation is a more rigorous process, it's more resource intensive, and there are advantages to doing outcome evaluation with selected interventions as well. That is why outcome evaluation is in the guidance, because it's important to build more capacity and more critical mass in that area across the country.

He said the message CDC wants to stress has to do with the relationship of outcome monitoring and outcome evaluation – the building block aspect, and they want to continue to point out the importance of knowing that a program has something well designed, and that it's actually being done the way in which a program thinks it is being done. Those are critical and necessary precursors to asking the questions about outcome. This is important to think about as people continue to consider which interventions may be appropriate for thinking about outcomes. Is it mature? Is it being implemented as designed? Is that likely to be the case throughout the period of data collection?

Discussion Summary:

Morning Session

- ☐ If there is a determination that they cannot do outcome evaluation, an inquiry was posed

as to whether they could use private funds to do so, and CDC funding to conduct outcome monitoring.

- ☐ Mr. Collins responded that it would seem that this would still not be appropriate, but he assured the participants that they would check on it and get back to them.

Afternoon Session

- ☐ No questions were posed during the afternoon session of the pyramid presentation.

Jeanine Ambrosio CDC Representative Role Play Exercise

Jeanine Ambrosio (acting as the CBO) and Charles Collins (acting as the Health Department) engaged in a role play exercise in which they received a cooperative agreement to conduct outcome evaluation [See copy of script]. Following their role play, the floor was opened for discussion.

Discussion Summary:

Morning Session

- ☐ An inquiry was posed as to whether participants should interpret the role play example in the context of the health department trying to select an intervention to satisfy the outcome evaluation requirement in the Guidance. Also, a question was posed as to how the health department would be providing technical assistance to CBOs which might be interested in doing that on their own. Charles Collins responded that the intent of the role play was to model some of the difficult questions that the health department would have to go through in terms of selecting an agency and an intervention for these types of evaluations. Also in the role play, in the end Jeanine Ambrosio (the CBO) asked for feedback. He thought that it would be common for programs to want technical assistance (for example, it may not be that CBOs are afraid of cost-effectiveness analysis per say, but they don't know how to do it and are seeking guidance).
- ☐ An inquiry was posed as to whether, based on the role play example, they would select outreach (e.g., Are the goals, objectives, and outcomes a program is trying to achieve appropriate for evaluation?). With the short encounters, they weren't talking about behavior change probably. Were they talking about behavioral determinants, or maybe having some effect on perceptions of risk? Charles Collins responded that if, in fact, the objective of the program was purely condom distribution, then the 2-second contact may

be appropriate. If the objective was to increase risk awareness, maybe an average of 5-minutes would do that. But if the CBO was really focusing on behavior change due to this intervention, she'll experience problems. She may not be able to get much behavior change for the amount of "dosage" that she is giving people in the street.

- ☐ A question was raised as to how follow-up, in the case of the role play scenario, would be conducted. The only way the participant saw that it could be done would be to ask the client to voluntarily agree to reveal their identity and to allow them to be contacted in some way on a post basis (1,2, or 3 months down the road by mail or phone) to administer some kind of risk assessment instrument. Another participant responded that the way they're not using names but are trying to do some type of evaluation where they're comparing people – they're using the "Stages of Change" model. On their evaluation forms they're asking people if it's the first time they've been spoken to by an outreach worker. Then they compare the evaluations where people have spoken to an evaluation worker more than once, to those evaluations where people have said they've only spoken to an outreach evaluation worker once. Their hope is that, over time, for those people who have spoken to an outreach worker more than once, they'll see a reduction in risk behaviors.
- ☐ A participant pointed out that studies showed that without an incentive (cash), follow-up would be difficult.
- ☐ A representative from Tennessee pointed out that the HIV prevention outreach workers are not the only people out collecting data. Sometimes clients they meet on the streets are confused as to which team they've talked to, so this can lead to the collection of incorrect information.
- ☐ A point was made that in the role play, only one strategy was used and that more should be planned for.
- ☐ An inquiry was posed as to whether doing a risk assessment was the intervention, or whether the few minutes that the client spends with the outreach worker is the intervention. If the outreach worker is making contact multiple times, that becomes the intervention. Charles Collins responded that that was what he was struggling with in the role play interview and why he wanted to see how the new workers were trained, because he wanted to learn exactly what the intervention was. Was it the assessment only? Was there some type of stage-based, tailored, or uniform message given? They didn't really get to the point in the role play of identifying what the intervention really was that took place in the 5-minute encounter.
- ☐ A suggestion was made that it would be nice to have a "cookbook" format of pre/post

measures. Charles Collins responded that Volume 2 of the *Guidance Supplemental Handbook* in the Outcome Monitoring Chapter, has some suggested measures for condom use and for injection drug use as a starting place. They're not the "gold standard" by any means, but they can be helpful. Gary Uhl added that they would have that solidified for directly funded CBOs. That should be finished in approximately a year and a half. He pointed out that participants who'd worked on projects such as the Special Projects for National Significance through HRSA and other funding mechanisms through the federal government where, in the initial stages of a cooperative agreement, those are laid out in the first year. All of the people who are funded collect whatever information they want to, but there is always a core set of common indicators. He said he found that lacking in the previous cooperative agreement for states. This will make it difficult to aggregate data. Moving this process in that direction is slow, but they're starting with the CBOs that were directly funded by CDC.

- A participant said that, given the need for correspondence between the content of the intervention and the measurement, if they are given core measures, that requires that their content match those core measures. With regard to the implication for CBOs, CDC was asked to comment on how they would deal with this, particularly since it's very difficult to find outcome measures that can go across interventions. Gary Uhl said CDC would pose a list of core questions that they think would be most appropriate to ask to determine changes in behavior. It would not necessarily be prescriptive for all interventions a CBO would fund, but a common set with which CDC is familiar and can provide suggestions about. There are lots of examples within the HIV/AIDS Division of consensus items and measures.
- A participant said that with regard to the Guidance and the recipe on how to do a plan, in the outcome evaluation section, it asks general questions. Obviously, the role play and discussions suggested that CDC wanted something much more detailed than that (e.g., sample size, how did you arrive at it, etc.). An inquiry was made as to what CDC planned to do in 2002 – when do they want the plans? Marlene Glassman responded that CDC really had no plans to review each jurisdictions proposal, sampling plan, or intervention. That doesn't mean they couldn't give some thought to doing that, but in all honesty, they had not planned on it. What they plan to do in the next week or so is to get a status of outcome evaluation and IRB involvement from each health department – not methodology, design sample, etc.
- A participant expressed concern about some of the "objective creep" they seemed to be hearing. She suggested that everyone go back and look at the purpose in the Guidance of why they are doing outcome evaluation. Her understanding was that it was to increase information availability about the effectiveness of different interventions – not to collect a national standard set of data. If that is, indeed, the point of the jurisdictions doing this,

then CDC not becoming incredibly involved in a critique of a proposal and design seems to be asking the jurisdictions to spend money on something that may not ultimately be considered of sufficient quality to enhance the knowledge base. Marlene Glassman responded that the participant was bringing up one of the major questions that distinguishes research from evaluation. CDC's argument has been (Chapter 7) that this is program evaluation.

- ❑ Participants urged CDC to take a more active role in working with the organizations and health departments to make sure their designs are solid, whether they end up calling it research or program evaluation. Marlene Glassman agreed that they should think about this issue. Charles Collins added that the issue of common indicators or measures would really be to assist health departments in not having to re-create them, but to know that there are some standard assessments. They're not saying that they should be required for all health departments, however. If this does go through as planned, there are regional trainings proposed for the fall.
- ❑ David Cotton noted that part of the tension in the room was what CDC often heard which is that some people are begging for structure, examples, and things to help them move along while others don't want CDC to come down with guidelines or limiting what they can do. Certainly, that creates a lot of tension and responding to both constituencies puts CDC in a difficult situation.

Afternoon Session

- ❑ One participant said they didn't know if they would get as specific as the health department did in the role play. He thought it was better to let the CBO just talk, because when the questions get too specific, the CBO is going to respond directly to those choices posed, causing the health department to miss out on key information that would give them a real assessment of the agency.
- ❑ Another participant pointed out that in many cases, the specifics weren't there, or was there fidelity to a particular model in the role play. For this CBO, things were very flexible in the field, they were doing whatever a particular client needed or whatever the new staff person was able to do, etc. Because of that, it's hard to pinpoint specifics and hold things to a particular model. Charles Collins responded that when they designed the role play, they were aiming for what the typical outreach program was like, and to show the struggles with trying to pin it down.
- ❑ A participant indicated that they went on visits to their CBOs, and one of the ways they got them to even think about this (they were typical of this role play) was to ask them, "If you were trying to get funding from someone, and had to prove that you were doing

something productive, what would you want to know that your agency was doing?” so that it put the onus back on the CBO to think about what they should know. This helped the CBOs create a list of what they wanted to know, because they thought they were doing it, but they weren’t sure. This gave them a great opportunity to set the work plan for their year’s agenda. They then went back to revise their goals and objectives to reflect putting these pieces in place, which made it an easier transition to evaluation.

- ❑ One participant expressed concern with sample size in the role play model, and inquired as to how an estimated 100 contacts per month that could range from 2 seconds to about 30 minutes, could produce a large enough sample size to do effective outcome evaluation. Charles Collins agreed, noting that it was one of the things that had happened in terms of the calls they’ve had from the states asking about appropriate sample size. One of the first things CDC ask the states is to think about the hypothesis and calculation of appropriate sample size. There needs to be enough sample size to answer the hypothesis.

Marcia Sass

Health Department Peer

New Jersey Department of Health and Senior Services

Marcia Sass said that as early as 1994, when the Community Planning Process was introduced in New Jersey, on day one, the consumers within the group demanded that not only should they do process evaluation, but also do outcome evaluation of their programs. So, evaluation has always been a priority in the New Jersey HIV Prevention Community Planning Group, and it was listed as a major priority in the first comprehensive plan in 1994. As soon as they had the opportunity to go for funding in 1995, they did so.

She said that either New Jersey was just lucky, or they were smart, but their populations and interventions have always been behaviorally based. In New Jersey, the leading mode of transmission has been injection drug use, followed by sexual transmission through infected partners. Their plan, populations, interventions, etc. have been based on addressing those risky sexual behaviors. So, they immediately submitted an application for supplemental funds to their 99094 agreement and received funding for a series of programs that came out of Community Planning. The recommended interventions/ services came out of research or effective programs, and the interventions were those that had been studied, and they were designed either to reduce injection drug use or risky sexual behaviors through behavioral interventions.

They launched a program in 1995. They developed a protocol, came up with a set of objectives for both process evaluation across the board on all of their agencies, and outcome evaluation on their agencies that identified prototype programs/ projects (For example: for injection drug users, for sex partners of injection drug users, and for at risk populations for sexual transmission such

as women, youth, and men who have sex with men). They developed a conceptual framework upon which their evaluation has been based. Because everything was being determined behaviorally, they decided that they could use one instrument, and that it could be developed so that if clients didn't participate in a particular behavior, they could skip out of that section.

Some of the things to keep in mind are that, at the state level, they did have evaluation capacity (e.g., people trained in outcome evaluation). The other thing, and this is probably consistent among jurisdictions, is that depending on where they are, the states procurement procedures, and even hiring make it extremely difficult to get the kinds of staff that might be needed to carry out these kinds of activities. They realized early on that they would need to work with a collaborator. Their choices were to get the resources within the division, by working with another department within the Department of Health, go to a sister state agency, or the worst fate – through the Department of Treasury and the procurement system because then Treasury makes the decision on who the evaluator is. They wanted to avoid that, so their choice was to go with the sister state agency. They created a legal agreement that spelled out the collaboration, the requirement for an advisory committee, and all of the deliverables for the particular collaborator.

Marcia Sass said they'd learned a lot as they moved along. They actually received the funding in September of 1995. It took about 18 months to get the agreement in place, and by the time the agreement was in place, it was about time to close it out and start over. That is what they did. They spent about \$5,000 to close out the first agreement and start over again working with the collaborator. She noted some of their difficulties:

- ☐ Countless hours have to be spent training collaborators. Even though well established in their communities and with many having plans, etc. the majority of their staffs are not inherently trained in how to deliver behavioral interventions. Training/re-training was necessary (and time consuming), and this led them to develop a training program, in which all of their staff and agencies are required to participate. The series includes 17 days of training, of which behavioral training is a major component. Still, turnover is rampant amongst both staff and interviewers. Training and retraining of interviewers has been necessary.
- ☐ Based on the amount of funding they received, they immediately had to scale back on the evaluation, and that included having to scale back on specific comparison groups. Since then, they learned that unless they'd done a full randomized control trial, there wouldn't be any comparison groups for the particular clients they have that would have given them any true association.
- ☐ There were no instruments to assist them to do the measurement. They combed the literature and built an instrument that would enable them to assess the outcomes. It took an enormous amount of time to put that together. To do the Stages of Change and

Transtheoretical models, most of them are scaled items with anywhere from 24 to 40 items. This would have taken three weeks to do a baseline if they'd incorporated all of the elements that were there.

- ☐ They were always working on buy-in at the state and community levels. They wanted the community level input, because they thought it was critical, but this was labor intensive. They needed to know if there was longer term improvement. So, they tried to study baseline and 6, 12, and 18 months after they'd enrolled in the program. After years of work, they have baseline and three follow ups documented.
- ☐ They also identified early on that they would need all of their instruments, anything associated with the instrumentation (e.g., the consent forms, hand cards, incentive receipt forms) in Spanish. That was an incredible challenge. It was difficult enough putting it together in English.
- ☐ They created training and coding manuals for each one of the interview schedules, which was also time consuming.

She noted some of their lessons learned/recommendations:

- ☐ Consider the real intents of doing evaluation, and how they translate into evaluation goals and objectives. New Jersey was specific, and they were never out to establish causality. They were looking to see whether they could identify *anything* that might be due to programs, and to build in program improvements.
- ☐ Buy-in is necessary at all levels. They've had more buy-in from their community folks than they have had at the state level. This is an unceasing activity.
- ☐ Evaluation is clearly dependent on the resources available. They can't do something like this without resources.
- ☐ They had enough time to assess their agencies by the time they went into the field. No matter how well their agencies were established in their communities, they are moving targets. There is churn, flux, chaos. They're stable one day and not the next, even if they're trying to implement the same intervention over time, particularly with behavioral interventions where so much of it is how an individual relates to a client. They also learned that even though their agencies were very well meaning, and they really wanted to do follow-up interviews, when it was time to do the follow-up, they weren't there and collaborating evaluato staff had to take over.
- ☐ Maintaining client confidentiality and privacy is a critical issue. They had to go to each

agency to ensure that there was space, etc. to make sure there would be privacy. The questions are fairly intrusive, so this required constant training and re-training to ensure that confidentiality/privacy issues were appropriately dealt with.

- ☐ What design will best fit is an issue in terms of evaluation goals and objectives. They have to determine what the emphasis should be (e.g., the process, process monitoring, evaluation, or impact – or all of those things).
- ☐ Not every evaluator can do program evaluation, and there aren't a lot of people who have any concept of how to go about doing this.
- ☐ Procurement procedures are a constant issue.
- ☐ They must realize that they can't please everyone all of the time. Their internal customers have been much more difficult to please than their agencies.

Discussion Summary:

Morning Session

- ☐ An inquiry was posed as to who pays for the 17 days of training. Marcia Sass indicated that the state health department pays for it.
- ☐ A number of participants were interested in obtaining copies of New Jersey's survey instruments. Marcia Sass responded that she would look into doing that. She said that part of the difficulty in getting the instrumentation done had to do with computer capacity, but there were periods of time that it was almost impossible to work on the instruments. Thus, it was unclear whether they would work on a website, etc.
- ☐ Noting that Rutgers was listed in Marcia Sass's materials, an inquiry was posed as to what role the scientific community plays in this, and what their role would be in the future (e.g., Would Rutgers publish the information?). Marcia Sass responded that the academic community provided a fair amount of information through the advisory committee, and both the department and Rutgers had opportunities to identify experts in sampling, research design, etc. The department also has very specific protocols in terms of what happens with the data and how it's presented, etc. They have in their work product statement specifically what the collaborator can do with the data. While the department will be working with Rutgers, they can't publish without the department's reviewing the information first and having the opportunity to have their names listed or provide disclaimers as to what's in there.

- With regard to Marcia Sass noting that she had more buy-in from the community than the health department, others said that was their experience as well. However, when it came down to actually doing the work, it was hard to get enough community involvement, participation, and dedication throughout the length of the process. Participants also wondered what New Jersey's mechanism was for reporting results back to the community. Marcia Sass was asked to comment more on that. Marcia Sass indicated that they have a wealth of data that covers many things. It takes a while to clean and analyze the data. They have engaged the agencies over time. About every six months, they have re-training retreats where they bring everybody together to discuss issues, problems, and give them specific training in various areas. These have been very helpful. They will soon conduct de-briefing sessions with each one of the agencies. This will be done one on one, and they will provide data and feedback.

Afternoon Session

- An inquiry was posed as to how long the implementation took. Marcia Sass responded that they started in 1998, and their delays came in the ability/inability to enroll sufficient samples sizes. This is one of the problems with trying to do comparison groups. They actually closed out baselines in December of 2000, and some of the agencies actually never made it to their targets.
- An inquiry was posed as to how large her staff is that's dedicated to this. Marcia Sass responded that her staff included herself, one person for the process monitoring/process evaluation side, she has two vacancies that she has a lot of difficulty filling, and she has a lot of outside support. Having a collaborator has been essential for her.
- A participant stressed that they should all work within their departments. Health department staff must buy in because this is a painstaking, long process.

Roger Myrick ***Health Department Peer*** ***California Department of Health***

Roger Myrick said that California is an interesting place to do HIV work because from the very beginning of the epidemic, they've had a lot of different constituents come together to put pressure on state legislators to provide funding for AIDS research. The program for which he currently works, which is affiliated with the University of California (the Universitywide AIDS Research Program – UARP), was formed in 1993 in response to activists, researchers, politicians, and educators across the state putting pressure on the state legislature to create some type of funding mechanism to provide dollars for AIDS research in a variety of areas (e.g, basic research, clinical research, and social behavioral research).

In 1995, things became more heated in terms of social behavioral research in California largely because activists were coming to the legislative table and pointing out the fact that UARP was not providing an equitable amount of dollars to social behavioral research. In response, the agency (UARP) designated and created a funding mechanism that provided dollars for a community collaborative prevention evaluation research that would fund partnerships between researchers in California, either at University of California research institutions or at non-profit research institutions, to partner with community prevention service providers to evaluate preventions or to study populations at risk who were receiving intervention services.

So, California is in a unique position because early in the epidemic, even before federal funders began to get involved in this issue, California took steps to create the basis for an infrastructure that would provide an on-going support basis for prevention evaluation research. In 1998, that program took a very important step in developing a partnership with the State of California, in the Department of Human Services State Office of AIDS. In the state office, they were preparing for the release of CDC's Evaluation Guidance. So, they were particularly interested in UARP's community collaborative program because they saw it as an opportunity to frame the entire Evaluation Guidance not only for CDC, but also for the State of California as a community collaborative response. In terms of community planning, that might be the most obvious first step, but also it is a first step in terms of developing process measures and ultimately moving to more outcome monitoring and outcome evaluation strategies.

One result of the partnership being formed in 1998 was that his position was created. It is a liaison position between the university system and the state health department system. It's an unusual type of job in that he's not supposed to be in either camp too much, and in both camps equally. While his home base is in the university system, the success of what he's doing is determined by the extent to which he can bridge those two communities that often have very different goals.

Given that context and framework, he discussed what they're doing in California to implement the Guidance and what they're doing in terms of outcome evaluation. The funds that they're receiving from CDC are primarily being used for infrastructure. The money they're using for outcome evaluation activities are coming from the State of California. So, they don't run into the kind of problems that have been brought up by the recent IRB development.

One of the first things that California did was to create, in the Office of AIDS Prevention Branch, a Prevention Research and Evaluation Section that was devoted solely to developing and implementing not only the CDC version of the Guidance, but also to addressing the statewide needs regarding evaluation. The second step was the collaboration with the University of California. Run out of the university president's offices, they are broader than any of the individual campuses. That's important because what they see with this collaboration is a collaboration between two statewide systems that then has impact on the more local health

jurisdiction level.

The next step that was critical was to get input from stakeholders which they did through a series of expert and stakeholder input meetings that included representatives from health departments, CBOs, and community planning groups from across the state, and prevention evaluation experts (across the state and nationwide). They called upon these people to assemble an action plan for key evaluation needs, concerns, and potential strategies.

The next step was to develop a state-specific guidance, largely based on the CDC Guidance. At the same time, they began to develop their web-based reporting system which will collect data on process implementation, and eventually on outcome monitoring. The system is being set up so that outcome monitoring will be an option in the future, but in order to get the system up and running in a timely manner, they're not able to have those fields immediately.

One of the things that is happening in California, and probably in many other states, is that local jurisdictions are currently collecting outcome monitoring data. As with their process implementation data, it's been coming in in narrative forms. It's impossible to do anything with, or even to ever really read. So, one of the things that the Evaluation Guidance and the web system is allowing them to do (and forcing them to do) is to systematize their data collection in terms of process and outcome monitoring.

In 1998, they also developed a plan for strategic technical assistance. Unfortunately, that plan didn't function well. They have now gone to new contractors in 2001. That is an on-going struggle, even in a state that has devoted many resources to this activity.

With regard to their outcome evaluation projects, even though all of them have a pre/post component, what they were primarily interested in in these project (because they were bringing together researchers and community providers) were projects that would be defined as outcome evaluation research – there is some type of comparison group. For all of the projects that they fund, they always require that the university or research institutions obtain IRB approval – even if it is a more simple pre/post design.

They essentially initiated a series of RFA processes that involve researchers, evaluators, health departments and CBOs statewide. The RFAs have been developed and coordinated through him and their partnerships with other stakeholders across the state, as well as nationally. Part of this effort involved statewide communication with all of their HIV prevention providers and researchers at the local level. So, it was a fairly major undertaking. They wanted to involve everybody in the process, so after each RFA was let out, they conducted statewide information sessions traveling around to different parts of the state to explain to people who might be interested in applying what the RFA was about, and the kinds of things that they needed to do in order to be successful.

A lot of this effort involved relationship building – identification of the players, who might be interested in this kinds of activity, who the potential collaborators would be, and getting the word out to them. Even more than that was the process of helping people make connections with people they might not have thought about. One of the things he did, primarily through large listservs, was to help link people in different geographic areas of the state with researchers who might be doing work either with similar populations or in the same geographic area. This seemed to be a particularly important piece depending upon the specific RFA.

Part of the preliminary work involved a lot of information dissemination in terms of literature about outcome evaluation, the CDC Guidance, the kinds of things that they were looking for in terms of outcome projects. So, they provided this general guidance up front, and they had a lot of characteristics in the RFA of collaborative activities that were required. Beyond that, a lot of the issues that came up earlier in terms of designating design, sample size, retention/recruitment strategies, etc. was left up to the researchers and the community partner to determine depending upon what worked best for them.

He thought there were advantages and disadvantages to both a more controlled effort and a less controlled one. It makes a lot of sense to let researchers be in charge of research design, but if too much freedom is allowed, then they get projects into the field that, even though they've had community input, once they're up and running then they start to deal with difficult issues like recruitment/retention that can fall apart. They've certainly had to deal with that.

One interesting thing about their review process is that their review committees are comprised of 50% academic researchers and 50% community service providers. Often in HIV research review committees, what they have is largely a research committee with community input. So, there will be 2 or 3 slots out of a 10-member panel who are determining how dollars are going to be awarded. Their set-up was very different because they had the 50/50 split, and they really see in their reviews that projects will come in with beautiful, elegant designs with which one could cut glass. However, they sounded very top-down and as though they didn't have the community input that they needed. So, even though they were beautiful and sound in terms of science, because the review committee was made up of 50% providers, they ranked as some of the lowest in the funding.

In terms of characteristics of the RFA, he likes to think of them as their principles of collaboration. They require documentation of these in the proposals that they receive from researchers. These projects were set up with dual principle investigators – one from the research organization and one from the community provider organization. Both people have equal amounts of power, and they get to determine budgets, so they didn't always get equal amounts of funding. What tended to happen is that the research organization got more funding.

However, in terms of determining the evaluation design, the use of the data from the evaluation,

how the evaluation will actually be managed and run were all collaborative decisions for which they require documentation in the proposals. He reviewed some of the types of evidence they required in the proposals to ensure that people weren't only *thinking* as an equal and collaborative team, but that they had to evidence that they were *functioning* in that way – even prior to funding.

Some of the additional requirements included the following:

- ☐ All projects must focus on high priority populations for the state (largely MSM in California – 75%)
- ☐ Documentation of implementation, outcomes, on-going collaborative activities, and on-going status of community infrastructure (staffing, stability, training).

He developed a dissemination plan to take information from these project and distribute that information to all health departments statewide. The components or elements that he is collecting from the projects include information on:

- ☐ Core elements of the interventions that were administered
- ☐ Core elements of the research project that were the foundation for the research activity (e.g., research protocol, instruments)
- ☐ Description of the necessary community organization infrastructure that had to be in place in order for the intervention to be successful

He is in the process of collecting this information now, and they should be able to disseminate that to the health departments across the state beginning in January, 2002. The difficulty lies in tailoring that information to make it relevant for other health departments, and ensuring that people aren't using other instruments for interventions that the instruments shouldn't be used for. That's something that they're in the process of dealing with.

They also have all of the grantees form a consortium that comes together twice per year to address issues that relate to community research needs. This has been a very interactive group, and one that's really helped move their process forward. They currently have 20 projects that are evaluating a variety of interventions, and a variety of populations. All of the projects that were funded from 1999 on have some type of control and/or comparison group. That's an essential part of the study. He said that it's unclear to him how successful the projects that started in 1999 are going to be. He thinks in a lot of ways they'll be very successful in terms of documenting some of the community information they need to get from these projects, but in terms of being able to say which piece of the intervention works with which type of population – he thinks they're going to be mixed results. The projects will probably be able to make some claims about which parts of the intervention worked, but he doesn't think their sample sizes in

the end are going to be large enough to be able to say which pieces worked best with which types of populations within the studies. In conclusion, he shared some materials with the participants.

Discussion Summary:

Morning Session

- ❑ A request was made for Roger Myrick to give them a sense of the typical time table and amount of funding for the outcome monitoring and outcome evaluation projects. Roger Myrick responded that these are typically 3-year projects. Funding varies depending upon what kind of infrastructure there is for the organizations because the health departments and the research infrastructure help support that to different degrees for different institutions. One project is being funded for \$500,000 for the entire 3-year project for both partners. That's a fairly small project. Another is funded for 3 years at \$900,000 and that's being shared between two partners. In that instance, that money is going to two organizations that have less infrastructure than other organizations. Their dollars were higher to compensate for that. It's about a \$2 million dollar a year program.
- ❑ An inquiry was posed as to whether Roger Myrick's collaborations always included a University of California school. He responded that it did not have to. They have state schools and PIs from non-profit research institutes as well. The thing they can't fund are privately funded researchers. They can come on as consultants, but they cannot be the main PI for the project.
- ❑ An inquiry was posed as to whether a health department could be a main PI on a project. Roger Myrick responded that they could, as could a person who is in an executive position at a CBO.
- ❑ An inquiry was posed as to whether, when considering the State of California, they excluded Los Angeles and San Francisco, given that these are considered separately by CDC. Roger Myrick responded that they did not exclude those cities. He said that even though Los Angeles and San Francisco are directly funded by the CDC, they also receive state funds, so they're very much a part of the state system both in terms of the university and the health department structures.
- ❑ An inquiry was posed to either or both peers as to whether they'd found behavior change. Marcia Sass responded that the primary role of their collaborator had been for the fielding and quality assurance, and not in terms of the analysis. They were trying to get on board, within the department, someone who was really skilled in that particular area. It's been an impossible situation. They now have a consultant who is skilled and trained

and will be working with the collaborator and the department so they will finally be producing some of the data within the next couple of months. Roger Myrick added that one thing they've learned is that this whole process has involved, and is leading them to, a total re-conceptualization (from both the research and health department perspectives) of what evaluation means in terms of service provision. They're having to re-educate both parties in order to bring people to common ground to make these efforts sustainable and to make them have any kind of long-term impact on the community or organizations in terms of research infrastructure. In thinking about TA, one of the things they've had to focus on is that they're not just teaching people to use reporting systems. They're orienting people to the activity of program evaluation and how it can help support their programs.

Afternoon Session

- ❖ Gary Uhl said that this programs seems to him to be very unusual, in-depth, and very interesting. However, he thought that it in no way supplanted what health departments are required to do regarding the Evaluation Guidance – this is above and beyond that. Roger Myrick responded that it was above and beyond Guidance requirements.
- ❖ An inquiry was posed as to how much the total award amounted to, or a projected cost, particularly given what appeared to be a very large staff and that these projects are long-term. Moreover, the epidemic is constantly changing.
- ❖ Roger Myrick responded that in terms of dollars, they're at the height of their funding, and are letting out about \$2 million per year in projects. That will go down. The projects typically run 3 years each. In terms of what the projects will be able to tell them, he said he thought that even with the difficulties that all of the projects have encountered with recruitment, they will be able to reflect that specific parts of interventions do or do not work to achieve certain kinds of change. He didn't think they'd end up with large enough sample sizes to be able to tell which changes work with which groups the best. Depending on the target population for the project, that may or may not be important. If there is a very homogenous group, this may be less important. But, in most cases, they have groups looking at at least 2 to 3 types of target populations that are related in some way. Regarding staffing, he really is the person who staffs it, he has one full-time research associate, and then he has pieces of people in the university and the health department.
- ❖ An inquiry was posed as to what percentage of the \$2 million came from CDC and what percentage came from the state legislature. Roger Myrick responded that the majority came from the state. In the beginning, the CDC money they got for supplemental projects jump-started the project and gave them some dollars for infrastructure.

However, in terms of the actual research projects themselves, depending on the nature of the project, probably 75% of the funding comes from the state legislature. He acknowledged that he needed to look for additional sources outside of that, such as soliciting other federal funders like NIH to support that. He stressed that it was critical not to look to one funder to provide resources for such a resource-intensive activity.

- ❖ A participant from Texas indicated that they and their evaluation partners are rapidly trying to do some extractions of the current literature to look at the core elements. He wondered if Roger Myrick had come up with a standardized taxonomy in terms of standardization of the core elements. Roger Myrick responded that they had not. With their dissemination plans, they have included the general materials that they want from the sites. As they begin to go around to the sites to collect these details, the sites will make these determinations for him.
- ❖ Several participants complimented the materials and program, but pointed out that in reality, in a number of rural states, programs would be lucky to get two trainers to actually follow the same protocol, and come up with a minimal number of clients. With that in mind, an inquiry was posed as to whether CDC had considered letting small to medium capacity states do a multi-site trial for one experiment in order to solve resource and other problems. Charles Collins responded that CDC would be very open to all types of creative activities, particularly if it turns out that they cannot use funds for outcome evaluation.

Gary Uhl
CDC Representative
Closing Exercise

Gary Uhl said the panel thought a good way to close the session would be to make a brief list of concrete, key issues which much be considered when conducting an outcome evaluation that they might want to tell their co-workers, or other people when they got back to their jurisdictions. The list included the following:

- ☐ What are the real intents and how do these translate into evaluation goal(s) and objectives?
- ☐ Whose buy-in do you need and how do you get it? Getting true buy-in for evaluation is an ongoing process at all levels. This includes providing and receiving feedback.
- ☐ Evaluation is dependent on what resources are available.
- ☐ Many of the agencies funded to implement HLV prevention interventions are “moving

targets.” Chaos is common. Not only are the clients transient, the staff are too. Program staff (and clients) you work with in the beginning may not be the same by the time you start your evaluation and/or finish it. Even if the stated interventions do not change, who is delivering them is likely to. A lot of flexibility is needed. If staff/community members are engaged to do the evaluation, the evaluator needs to be able to “jump-in” and continue on with the activities when they become overwhelmed.

- ☐ What language(s) are needed in the data collection instruments must be considered.
- ☐ How client confidentiality and privacy can best be maintained must be addressed on an agency by agency basis.
- ☐ What design will best fit the evaluation goals and objectives?
- ☐ Determine where the emphasis should be on inputs/process, outputs/impact or all?
- ☐ How do you select an evaluator? / Who should do it and how should you relate? A collaboration often is needed. Don’t assume that all evaluators are capable of accomplishing outcome evaluation.
- ☐ Determine how your jurisdiction’s procurement procedures work to provide you with the greatest flexibility in selecting an evaluator.
- ☐ Realize that you can’t please everyone all the time, in particular those internal customers who need instant gratification (immediate data). Outcome evaluation can take time. Changes in the design and other evaluation activities are often needed. Stick with it.

End of Summary Proceedings

